DOCUMENT RESUME

ED 126 427 CS 002 780

AUTHOR Horn, Thomas D.; Hernandez, Carmela
TITLE Standardized Testing and the Spanish-Speaking Minorities: Looking Beyond the Correlations.
PUB DATE 76
NOTE 16p.

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Aptitude Tests; Bilingual Students; Elementary Education; *English (Second Language); *Reading Readiness; *Reading Readiness Tests; Reading Research; Research Methodology; Research Reviews (Publications); *Spanish Speaking; *Standardized Tests

ABSTRACT

This paper contains a critical evaluation of the research conducted by O.L. Davis and Carl Personke on the use of reading readiness tests in English and Spanish for Spanish speaking elementary school pupils. Davis and Personke indicated that, when Spanish speaking first graders were tested in both English and Spanish, most of the differences were not significant. The authors of this critique argue that Davis and Personke misinterpreted their data and that the result has been a continued misuse of standardized tests for Spanish speaking pupils. One of the objections was that insufficient attention was given to language dominance factors and to the identification of pupil characteristics other than language. Another problem was that the nature of the population distributions contributing to the correlations between predictor and criterion measures had not been carefully analyzed. (MKM)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED126427

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Thomas D. Horn and
Carmela Hernandez
The University of Texas
at Austin 78712

STANDARDIZED TESTING AND THE SPANISH-SPEAKING MINORITIES: LOOKING BEYOND THE CORRELATIONS

In the decade since the inauguration of the U. S. Office of Education First Grade Reading Studies in 1964 (The Reading Teacher, May-October, 1966) questions have been raised concerning the assessment of minority group children (Oakland, 1973), special education placement (Ross, Young and Cohen, 1971), and the use of available reading readiness and achievement tests with children who come from a Spanish-speaking background (Horn, 1966). More recently, succinct statements have been made concerning the implications for assessment of language characteristics for Black children (Bartell, Grill, and Bryen, 1973) and Mexican-American children (Matluck and Mace, 1973). In addition, reviews of major theories of language development and their implications for reading instruction suggest that reading research and methodologies are on the threshold of major changes (DiVesta and Palermo, 1974).

Despite an increased tempo of interest in the measurement of reading performance (Blanton, Farr and Tuinman, 1974) and the evaluation of other language arts (Fagani, Cooper and Jensen, 1975), the way that tests are being used in our schools continues to generate criticism (Goodman, 1975), particularly in urban schools (MacGinitie, 1973). Of prime concern are test validity (Allen, 1974), test fairness (Flaugher, 1974), tests and the disadvantaged

CS 002 780

(Fitzgibbon, 1974), and the use of tests in English with Spanish-speaking elementary school pupils (Davis and Personke, 1968; Personke and Davis, 1969).

The remainder of this article is a critique of the Personke and Davis studies and of the misinterpretation of their findings resulting in the continuing misuse of standardized tests with Spanish-speaking school pupils.

Statement # 1 (Personke and Davis, 1969: pp. 79-80)

> "When Spanish-speaking first-graders in four classrooms were tested in both English and Spanish, most of the differences were not significant. These findings suggest that the English administration of the Metropolitan Readiness Tests probably did not result in inadequate assessment of, or testing bias against, Spanish-speaking children, at least as far as the language parameter was concerned."

Responses to Statement # 1

(1) Without any oral language assessment to determine the language dominance of the subjects, it is not known to what extent the language factors contributed to the scores, i.e., monolingual English subjects' scores on the Spanish administration and monolingual Spanish subjects' scores on the English administration of the Metropolitan Readiness Test (MRT);

(2) the MRT does not sample for established language dominance in the standardization population;

(3) since no description of instructional practices in reading is provided, the question of the predictive validity of

the English and Spanish administrations of the MRT must remain unanswered; the fact that the MRT scores from the Spanish administration were not correlated with a Spanish version of the Metropolitan Achievement Test (MAT), other than none was available, possibly reflects an instructional situation where beginning reading instruction was in English only;

(4) the issue of whether or not translated tests (paraphrased is probably a more accurate term) are appropriate for native speakers of Spanish has not been resolved;

(5) a different kind of statement is proposed by this author as to what the findings suggest and gives quite another perspective: The English administration of the Metropolitan Readiness Tests probably did not result in any more inadequate assessment of, or testing bias against, Spanish-speaking children than did the Spanish administration of the MRT; that is, one was as inadequate as the other;

(6) any statement to the effect that any test in a language other than that of the child being tested is not biased is, on the face of it, absurd;

(7) Table 1 is based upon Phase 1 of the Personke and Davis study and underscores the reasons why researchers must look beyond computed correlations. In the case of Group A, these children were judged deficient in the English language and placed in pre-first or "readiness classes." The thirty five children assigned to regular first grade classes (Group B) were apparently

TABLE 1

Davis and Personke, J Ed Meas, Fall, 1968, p. 232

MRT Group A (N=53) "Pre-First" Sections		MRT Group B (N=35) Regular First Gr. Sections		MRT Total Group (N=88)	
English		# English		English	
Mean	39.93	Mean	49.86	Mean	44.67
SD	13.65	SD	13.44	SD	14.43
%ile	23	%ile	40	%ile	31
Spanish		Spanish		Spanish	
Mean	33.98	Mean	55.10	Mean	44.06
SD	12.82	SD	12.50	SD	16.48
*%ile	(15)	*%ile	(51)	*%ile	(29)
r	.76		.73		.63

*parentheses indicate %ile based on English administration distribution, MRT Manual of Directions, p. 11.

#difference favoring Group B English mean over Group A English mean significant ($t=3.44$; $p<.01$). Group A vs. Group B mean difference for Spanish version apparently not tested for significance, but is about twice that for the English version.

judged ready to begin regular reading instruction. Examination of the data strongly suggests that Group A and Group B represent two different pupil populations as supported by the following:

(a) The difference favoring the Group B English mean over the Group A English mean was statistically significant (t=3.44; probability less than .01); the Group A mean of 39.93 represents the 23rd percentile while the Group B mean of 49.86 represents the 40th percentile;

(b) although no t-test was apparently computed for the difference favoring the Group B Spanish mean (55.10) over the Group A Spanish mean (33.98), this difference is over twice that for the English version. In addition, on the possibly questionable assumption that the percentiles based on the English administration distribution might be comparable to the Spanish administration, the Group A mean would fall at the 15th percentile while the Group B mean would fall at the 51st percentile. In any case, pupils assigned to Group B by virtue of being judged competent in English also scored much higher than Group A on the Spanish version. It is an interesting question as to why the Group A pupils scored lower on the Spanish version than they did on the English while the reverse is true for the Group B students.

(c) When Group A and B are combined the nature of the distributions is obscured so that the English (44.67) and Spanish (44.06) means are almost identical; the correlation coefficient, though somewhat lower (r=.63), is still comparable

to the correlations between the English and Spanish versions for Group A (r=.76) and Group B (r=.73).

Statement # 2 (Personke and Davis, 1969: p. 81)

> "The number of subjects reported here is half of the number reported in Phase I. The loss was not accidental. The two 'readiness' classes had been introduced to reading so late in the year that it was impossible to administer a reading test at the end of the year . . . These children had very low scores on the Metropolitan Readiness Tests and did not learn to read."

Responses to Statement # 2

(1) In the words of Personke and Davis (1969, p. 84), "It is not enough to note that a test is a valid predictor of success in reading if the prediction for a large group of children is failure"; this suggests the need to look beyond the correlations;

(2) in Horn's 1966 study, any student who successfully guessed the correct answer on three of the sixteen MRT Word Meaning items fell at the project mean (2.99; SD=2.01; National Mean = 8.67; SD=8.67);

(3) also in Horn's 1966 study, when 98% of the project sample fell in the two lowest MRT categories (D, low normal; and E, low) for the Word Meaning subtest, there is little discrimination between levels of ability; this suggests questionable test validity for this population; similarly

questionable test validity appears to exist for the Personke and Davis population;

(4) any meaningful pupil perception of testing task was unlikely in Horn's 1966 sample when about 28% (164 of 584) scored 0 or 1 on the Word Meaning subtest; similar information is not available for the Personke and Davis study.

Statement # 3 (Personke and Davis, 1969: p. 82)

"The Subtests of Word Meaning and Listening, of the Metropolitan Readiness Tests, were not highly related to reading success."

Response to Statement # 3

(1) Rather than use the euphemism "not highly related" with reference to the Subtests of Word Meaning and Listening in both the Spanish and English administrations, the r's shown in Table 2 below warrant the following:

(a) There was no relationship between the MRT Word Meaning and MAT Word Knowledge, Word Discrimination and Reading on the English and Spanish administrations respectively;

(b) there was no relationship between the MRT Listening subtest and MAT Word Knowledge, Word Discrimination and Reading on the Spanish administration;

(c) the relationship between the MRT Listening subtest for the English administration was low re: the MAT Word Knowledge and Word Discrimination; the correlation of .33 with the MAT Reading subtest is significant at the .05 level of competence;

TABLE 2

	MRT Subtest 1, Word Meaning		MRT Subtest 2, Listening	
	English	Spanish	English	Spanish
MAT Word Knowledge	-.01	.00	.28	.02
MAT Word Discrimination	.04	.03	.18	.05
MAT Reading	-.03	.16	.33*	.08

*Significant at the .05 level

(d) the foregoing instances of no relationship very likely reflect a mixture of Spanish-dominant and English-dominant subjects in the sample; language dominance should be established prior to instruction.

Statement # 4 (Personke and Davis, 1969: pp. 82-83)

> "These findings indicate that the Metropolitan Readiness Tests are useful in predicting certain reading-related achievements for Spanish-speaking first-grade pupils. Administration of the Metropolitan Readiness Tests in Standard English and colloquial Spanish seem to yield similar predictions. Consequently, administration of the test in English rather than in the pupils' native Spanish apparently does not result in test bias. The evidence does not support moves to question the general practice of administering tests in English to Spanish-speaking school entrants."

Responses to Statement # 4

(1) Table 3 pulls together in tabular form the narrative presentation of data found on p. 82 of Personke and Davis and compels the following observations concerning how useful the MRT is in predicting "certain reading-related achievements for Spanish-speaking first-grade pupils":

(a) Relationships between the total MRT score and MAT Word Knowledge were significant at the .01 level of competence for both the English administration (r=.87) and the Spanish MRT administration (r=.61); in fact, all correlations are significant at either the .05 or .01 level of competence as reported by Personke and Davis and presented in Table 3;

TABLE 3

PERSONKE-DAVIS SIGNIFICANT MRT PREDICTORS OF MAT PERFORMANCE FAVORING:

English MRT Administration		#Spanish MRT Administration	
Total MRT Score - MAT word knowledge	t=6.45; p<.01 ##(r's=.87†; .61†)	Copying subtest - word knowledge	t=2.34; p<.05 ##(r's=.64†; .39*)
Alphabet subtest - word knowledge	t=4.00; p<.01 (r's=.78†; .38*)	Copying subtest - word discrimination	t=2.18; p<.05 (r's=.69†; .47†)
Alphabet subtest - word discrimination	t=3.13; p<.01 (r's=.77†; .45†)	Copying subtest - reading	t=7.34; p<.01 (r's=.60†; .37*)
Alphabet subtest - reading	t=3.22; p<.01 (r's=.65†; .32*)		

†Significant at the .01 level
*Significant at the .05 level

#Identified (Personke and Davis, p. 82) as "colloquial Spanish."

##Correlation, given first is for English administration in column one; the Spanish administration in column two.

(b) the same number of subtests, three each, favor the English and Spanish administration of the MRT;

(c) for the correlation that ordinarily would be considered the most significant, i.e., between the total MRT score and the MAT Reading, none was significant for either the English or Spanish administration;

(2) an immediate question to be raised concerning the usefulness of MRTs for prediction purposes has to do with the economic and human defensibility of subjecting linguistically distinct children to a testing task in which approximately half failed (Personke and Davis, 1969: p. 81); although no comment is made concerning the psychological impact of failure on young children, a statement is made earlier concerning the impact on correlation computations; "it is doubtful that the inclusion of data on the reading achievement of the readiness classes would have effected the coefficients of correlation in any but a positive way, since these children had very low scores on the Metropolitan Readiness Tests and did not learn to read" (Personke and Davis, 1969: p. 81); the time, expense and negative psychological impact of such testing is open to serious question;

(3) the statement to the effect that the administration of tests in English rather than in the pupils' native Spanish did not result in test bias, as mentioned earlier, represents a prima facie absurdity; test bias results when <u>any</u> child is tested in a language other than the one which is dominant for him;

(4) the use of <u>any</u> currently published standardized readiness tests with <u>Spanish-dominant</u> school beginners is highly questionable,

especially in the southwest; in addition to cultural biases in test items (e.g., "mark the toboggan" verses "mark the tamale."); tests such as the MRT fail to include proportional numbers of linguistically different (or distinct, if you prefer) pupils in the standardization sample; even with the inclusion of a limited number of so-called Spanish-surname pupils, this is meaningless without oral language assessment to determine language dominance;

(5) there would be few objections to administering tests in English to Spanish-speaking school entrants if, in fact, such pupils are true bilinguals, i.e., as at home with English as with Spanish; for whatever reason, the number of Spanish-surname school beginners who are monolingual in English (or possibly more to the point, those who decline to acknowledge Spanish) is noticeably higher than just a few years ago.

Conclusions

When applying the results of reading research to the classroom, a critical review of the research and research methodology is basic; in this case:

(1) lack of attention to language dominance factors and the identification of differing pupil characteristics other than language constitute significant weaknesses in the study;

(2) instead of accepting at face value the validity of statistically significant correlations between predictor and criterion measures, the nature of the population distributions contributing to those correlations needs to be carefully analyzed, particularly when distributions appear to be bi-modal or nearly so; and

(3) unless the foregoing are made components of the research design and data analysis, the result is a continuing inappropriate

use of MRTs with Spanish-dominant school beginners.

BIBLIOGRAPHY

Allen, Virginia F. "What Does a Reading Test Test?" Temple University College of Education Triple "T" Project Monograph Series, 1. Philadelphia, Pennsylvania: Temple University, 1974, 1-21.

Bartel, Nettie R., J. Jeffrey Grill, and Diane N. Bryen. "Language Characteristics of Black Children: Implications for Assessment." Journal of School Psychology, 11 (December, 1973), 35-64.

Blanton, William E., Roger Farr, and J. Jaap Tuinman. Measuring Reading Performance. Newark, Delaware: International Reading Association, 1974.

Davis, O. L., Jr., and Carl R. Personke, Jr. "Effects of Administering the Metropolitan Readiness Test in English and Spanish to Spanish-speaking School Entrants." Journal of Educational Measurement, 5 (Fall, 1968), 231-34.

Di Vesta, Francis J., and David S. Palermo. "Language Development." Review of Research in Education, Vol. 2. Itasca, Illinois: F. E. Peacock, 1974, 55-107.

Fagan, William T., Charles R. Cooper, and Julie M. Jensen. Measures for Research and Evaluation in the English Language Arts. Urbana, Illinois: ERIC/RCS and National Council of Teachers of English, 1975.

Fitzgibbon, Thomas J. "Reading Tests and the Disadvantaged." In Measuring Reading Performance, pp. 15-33. Edited by William E. Blanton, Roger Farr, and J. Jaap Tuinman. Newark, Delaware: International Reading Association, 1974.

Flaugher, Ronald L. "The New Definitions of Test Fairness in Selection: Developments and Implications." Educational Researcher, 3 (October, 1974), 13-16.

Goodman, Kenneth S. "Do You Have to be Smart to Read? Do You Have to Read to be Smart?" The Reading Teacher, 28 (April, 1975), 625-32.

Horn, Thomas D. "Three Methods of Developing Reading Readiness in Spanish-Speaking Children in First Grade." The Reading Teacher, 20 (October, 1966), 38-42.

MacGinitie, Walter H. "Testing Reading Acheivement in Urban Schools." The Reading Teacher, 27 (October, 1973), 13-21.

Matluck, Joseph H., and Betty J. Mace. "Language Characteristics of Mexican-American Children: Implications for Assessment." Journal of School Psychology, 11 (December, 1973), 365-86.

Oakland, Thomas D. "Assessing Minority Group Children." Journal of School Psychology, 11 (December, 1973), 294-303.

Personke, Carl R., Jr., and O. L. Davis, Jr. "Predictive Validity of English and Spanish Versions of a Readiness Test." Elementary School Journal, 70 (November, 1969), 79-85.

Ross, Sterling L., Henry G. DeYoung, and Julius S. Cohen. "Confrontation: Special Education Placement and the Law." Exceptional Children, 38 (September, 1971), 5-12.

The Reading Teacher, 19 (May, 1966) and 20 (October, 1966).